

# Enhancing source separation quality via optimal sensor placement in noisy environments

Mohammad SADEGHI<sup>a,b,\*</sup>, Bertrand RIVET<sup>a</sup>, Massoud BABAIE-ZADEH<sup>b</sup>

<sup>a</sup>Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, 38000, France

<sup>b</sup>Department of Electrical Engineering, Sharif University of Technology, Tehran, 1458889694, Iran

---

## Abstract

The paper aims to bridge a part of the gap between source separation and sensor placement studies by addressing a novel problem: “Predicting optimal sensor placement in noisy environments to improve source separation quality”. The structural information required for optimal sensor placement is modeled as the spatial distribution of source signal gains and the spatial correlation of noise. The sensor positions are predicted by optimizing two criteria as measures of separation quality, and a gradient-based global optimization method is developed to efficiently address this optimization problem. Numerical results exhibit superior performance when compared with classical sensor placement methodologies based on mutual information, underscoring the critical role of sensor placement in source separation with noisy sensor measurements. The proposed method is applied to actual electroencephalography (EEG) data to separate the P300 source components in a brain-computer interface (BCI) application. The results show that when the sensor positions are chosen using the proposed method, to reach a certain level of spelling accuracy, fewer sensors are required compared with standard sensor locations.

*Keywords:* Source separation, sensor placement, statistical signal processing

---

## 1. Introduction

Determining optimal sensor positions to maximize the quality of a data measurement system is a crucial consideration preceding data analysis and processing, forming a part of the experiment design. This concern has been investigated in various applications where sensors are responsible for acquiring information, such as wireless sensor networks [1, 2], structural health monitoring [3] or source

---

\*Corresponding author

*Email addresses:* mo.sadeghi@sharif.edu (Mohammad SADEGHI),  
bertrand.rivet@gipsa-lab.grenoble-inp.fr (Bertrand RIVET), mbzadeh@yahoo.com  
(Massoud BABAIE-ZADEH)

localization [4, 5]. This paper focuses on the novel application of optimal sensor location prediction for source separation.

The objective of a source separation problem [6] is to estimate the latent source signals from their mixtures recorded by the sensors. Although source separation has been extensively studied from the late 80's [6] to propose the best algorithm to separate or extract the sources, very few has been done to study the impact of sensor location onto the quality of the estimated sources. However, the position of the sensors has directly an impact on the mixing model and thus on the recorded signals. For instance, in a linear source mixing model, each sensor measures a signal that is a linear combination of the sources, and the coefficients of this combination depend directly on the sensor position. In the absence of additive noise, the coefficients, and consequently the sensor positions, do not affect the quality of the source separation due to the existence of equivariant blind source separation (BSS) methods that provide performance independent from the mixing coefficients [7, 8]. This might explain why sensor placement for source separation has not already been thoroughly investigated in the literature. However, as this paper demonstrates, optimal sensor placement becomes a crucial consideration for source separation in the presence of additive noise in the measurements.

There have been a few studies addressing the sensor placement problem for source separation in acoustic environments, when sources are convolutively mixed, without accounting for the impact of additive noise. In [9], the sensitivity of the unmixing system is analyzed in relation to errors in both the mixing system and observations. In [10], efforts are made to optimize microphone locations to improve the quality of the desired separated speech. However, it assumes that the intended speech source signals are known, limiting thus the generality of this solution. The current paper considers noise as a crucial parameter in the problem as it is unavoidable in practical situations, targeting thus most applications of source separation. An application example is electroencephalography (EEG) experiments, in which signals recorded by electrodes consist of a mixture of sources of interest and interfering noise from ongoing brain activities and artifacts. Consequently, optimal electrode placement on the scalp is crucial for achieving the best quality of the estimated sources.

It is important to note that the sensor placement problem is fundamentally distinct from the sensor selection problem, as explored in studies such as [11, 12]. In sensor selection, data from pre-existing sensors is available, and the goal is to select some of them with the most useful information. However, in sensor placement, the objective is to find the best positions to place a few sensors without accessing their data before deployment.

A related topic to the current work is sensor placement for target localization, which has been widely studied in the literature [13, 14, 15]. These studies optimize sensor locations to maximize the accuracy of localizing one or multiple targets. The key differences between these works and the current problem are as follows: first, sensor measurements in previous studies typically involve time-of-arrival (TOA), angle-of-arrival (AOA), or received-signal-strength (RSS) metrics, rather than a mixture of source signals. Second, the goal in these stud-

ies is to estimate the target location, rather than separating the latent source signals. Despite some similarities, the key differences between source separation and target localization make them two distinct problems.

Optimal sensor placement for extracting a single source from noisy measurements has been studied in [16, 17]. To address the optimal sensor placement problem for separating mixed sources from noisy sensor measurements, the current paper uses the signal-to-interference-plus-noise ratio (SINR) and the mean squared error (MSE) measures as the optimization criteria, combining them with the stochastic model of the source gains, leading to a stochastic optimization problem. To tackle this optimization problem, a gradient based method is developed, significantly enhancing computational efficiency compared with the grid search method employed in previous studies.

Parts of this work have already been presented in the conference paper [18]. However, the current paper provides a more detailed analysis of numerical results compared with that conference paper, along with the inclusion of results from applying the proposed method to actual P300 brain-computer interface (BCI) data. Moreover, it introduces new contributions on the optimization aspect by incorporating the gradient of the objective function. Additionally, the paper introduces the MSE sensor placement criterion for source separation.

The paper is organized as follows. In Section 2, the source separation model and the proposed criteria are presented. Section 3 discusses the optimization approaches for the problem. Finally, in Section 4, our method is numerically studied, and the results of applying the method to actual EEG data are included.

## 2. Sensor placement for source separation

This section begins by explaining the linear source mixing and separating model. Then, two criteria for optimal source separation are presented. Finally, the discussion revolves around the reasoning and the procedure behind the modelling of the spatial gains using the Gaussian process (GP) model.

### 2.1. The mixing model of the sources

Consider  $P$  independent latent sources,  $s_1(t), s_2(t), \dots, s_P(t)$ . In order to measure the mixed source signals, the sensors are to be placed in a  $D$ -dimensional space,  $\mathcal{X} \subset \mathbb{R}^D$ . The signal attenuation from the  $p$ -th source to the sensor located at the coordinates  $\mathbf{x}$  is denoted as  $a_p(\mathbf{x})$ , and it is referred as the spatial gain of the  $p$ -th source at  $\mathbf{x}$ . Assuming an instantaneous linear mixing model of the sources in the presence of additive noise, the measured signal by the sensor located at the coordinates  $\mathbf{x}$  is expressed as

$$y(\mathbf{x}, t) = \sum_{p=1}^P a_p(\mathbf{x}) s_p(t) + n(\mathbf{x}, t), \quad (1)$$

where  $n(\mathbf{x}, t)$  represents the noise signal at the sensor position. Note that in the mixing model, the propagation delay between the sources and the sensors is

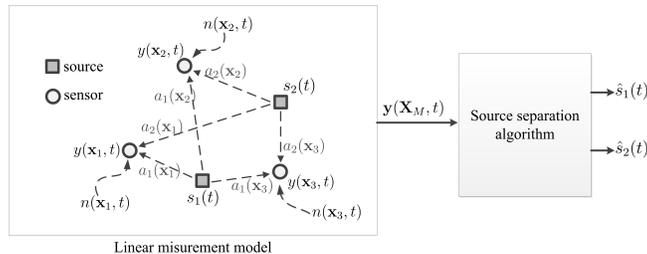


Figure 1: Linear measurement and source separation model, with 2 sources and 3 sensors.

assumed to be negligible. This assumption is valid for electrical signal propagation, such as in electroencephalography (EEG) source separation experiments [19, 20]. However, in the case of the acoustic environment, a convolutive mixing model should be taken into account.

The noise is assumed to be independent of the sources and may exhibit spatial correlation. The source signals are assumed to be independent of each other and have zero means and unit variances. This variance assumption does not limit the generality of the problem, as the power of the sources can be embedded in their spatial gains.

Consider a set of  $M$  sensors located at the positions  $\mathbf{X}_M \triangleq [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]$ . The vector of measured signals by these sensors at time  $t$  is denoted as  $\mathbf{y}(\mathbf{X}_M, t) = [y(\mathbf{x}_1, t), y(\mathbf{x}_2, t), \dots, y(\mathbf{x}_M, t)]^T$ , and can be expressed as

$$\mathbf{y}(\mathbf{X}_M, t) = \sum_{p=1}^P \mathbf{a}_p(\mathbf{X}_M) s_p(t) + \mathbf{n}(\mathbf{X}_M, t), \quad (2)$$

where the vector  $\mathbf{a}_p(\mathbf{X}_M) \triangleq [a_p(\mathbf{x}_1), a_p(\mathbf{x}_2), \dots, a_p(\mathbf{x}_M)]^T$  contains the spatial gains of the  $p$ -th source, and  $\mathbf{n}(\mathbf{X}_M, t) \triangleq [n(\mathbf{x}_1, t), n(\mathbf{x}_2, t), \dots, n(\mathbf{x}_M, t)]^T$  denotes the vector of the additive noise at the sensor positions  $\mathbf{X}_M$ . Fig. 1 illustrates the problem setup with  $M = 3$  sensors measuring the propagated signals of  $P = 2$  sources.

## 2.2. Criteria for sensor placement

Here, by assuming that the sources are separated linearly, two criteria for sensor placement are obtained. Note that the goal is not to optimize the separation algorithm, but to optimize the position of the sensors given a separation method. The first criterion is based on the SINR of the separated signals, and the second one is based on the MSE of the separation. These criteria are two measures of the quality of the source separation.

### 2.2.1. SINR criterion

In a linear estimation scheme, the  $l$ -th source is estimated by the inner product of a vector  $\mathbf{f}_l \in \mathbb{R}^M$  with the vector of measurements  $\mathbf{y}(\mathbf{X}_M, t)$ , which

can be expressed as

$$\hat{s}_l(t) = \mathbf{f}_l^T \mathbf{y}(\mathbf{X}_M, t) = \sum_{p=1}^P \mathbf{f}_l^T \mathbf{a}_p(\mathbf{X}_M) s_p(t) + \mathbf{f}_l^T \mathbf{n}(\mathbf{X}_M, t). \quad (3)$$

The estimated signal contains terms corresponding to the source of interest ( $\mathbf{f}_l^T \mathbf{a}_l(\mathbf{X}_M) s_p(t)$ ), the remaining sources acting as interferences ( $(\sum_{p=1, p \neq l}^P \mathbf{f}_l^T \mathbf{a}_p(\mathbf{X}_M) s_p(t))$ ), and the noise ( $\mathbf{f}_l^T \mathbf{n}(\mathbf{X}_M, t)$ ). The SINR is used as a measure for the quality of the  $l$ -th separated source, and can be written as

$$\text{SINR}_l(\mathbf{f}_l; \mathbf{X}_M) = \frac{\mathbb{E} \{ (\mathbf{f}_l^T \mathbf{a}_l(\mathbf{X}_M) s_l(t))^2 \}}{\mathbb{E} \left\{ \left( \sum_{p=1, p \neq l}^P \mathbf{f}_l^T \mathbf{a}_p(\mathbf{X}_M) s_p(t) + \mathbf{f}_l^T \mathbf{n}(\mathbf{X}_M, t) \right)^2 \right\}}. \quad (4)$$

The noise covariance matrix is defined as  $\mathbf{C}^n(\mathbf{X}_M, \mathbf{X}_M) \triangleq \mathbb{E} \{ \mathbf{n}(\mathbf{X}_M, t) \mathbf{n}(\mathbf{X}_M, t)^T \}$  and assumed to be known, denoted as  $\mathbf{C}_{MM}^n$  for simplicity throughout the paper. Using it and the statistics of the source signals, the SINR can be simplified to

$$\text{SINR}_l(\mathbf{f}_l; \mathbf{X}_M) = \frac{\mathbf{f}_l^T \mathbf{a}_l(\mathbf{X}_M) \mathbf{a}_l(\mathbf{X}_M)^T \mathbf{f}_l}{\mathbf{f}_l^T \left( \sum_{p=1, p \neq l}^P \mathbf{a}_p(\mathbf{X}_M) \mathbf{a}_p(\mathbf{X}_M)^T + \mathbf{C}_{MM}^n \right) \mathbf{f}_l}. \quad (5)$$

Our objective is to use the maximum achievable SINR with the sensor positions as the sensor placement criterion. Thus,  $\mathbf{f}_l$  is chosen to maximize the SINR in (5). The optimum  $\mathbf{f}_l$  is given by [21]

$$\mathbf{f}_l^* = \left( \sum_{p=1, p \neq l}^P \mathbf{a}_p(\mathbf{X}_M) \mathbf{a}_p(\mathbf{X}_M)^T + \mathbf{C}_{MM}^n \right)^{-1} \mathbf{a}_l(\mathbf{X}_M). \quad (6)$$

Note that the noise covariance matrix is assumed to be full rank, ensuring the matrix is invertible in (6). By substituting (6) into (5), the maximum achievable SINR for the  $l$ -th source using a linear estimation is given by

$$\text{SINR}_l(\mathbf{f}_l^*; \mathbf{X}_M) = \mathbf{a}_l(\mathbf{X}_M)^T \left( \sum_{p=1, p \neq l}^P \mathbf{a}_p(\mathbf{X}_M) \mathbf{a}_p(\mathbf{X}_M)^T + \mathbf{C}_{MM}^n \right)^{-1} \mathbf{a}_l(\mathbf{X}_M), \quad (7)$$

which is a function of the sensor positions. To consider the separation of all the sources, the sum of the SINRs of the sources is chosen as a criterion for sensor placement:

$$J_{\text{SINR}} = \sum_{l=1}^P \text{SINR}_l(\mathbf{f}_l^*; \mathbf{X}_M). \quad (8)$$

Note that this criterion can be modified to fit cases where only the separation of a subset of the sources is desired. In such cases, only the SINR of the desired sources will be included in the summation:

$$J_{\text{SINR}}(\mathcal{L}) = \sum_{l \in \mathcal{L}} \text{SINR}_l(\mathbf{f}_l^*; \mathbf{X}_M), \quad (9)$$

where  $\mathcal{L} \subset \{1, 2, \dots, P\}$  is the subset of desired sources.

### 2.2.2. MSE Criterion

Let us consider an alternative form of the mixing model (2) given by

$$\mathbf{y}(\mathbf{X}_M, t) = \mathbf{A}(\mathbf{X}_M)\mathbf{s}(t) + \mathbf{n}(\mathbf{X}_M, t), \quad (10)$$

where  $\mathbf{A}(\mathbf{X}_M) \triangleq [\mathbf{a}_1(\mathbf{X}_M), \mathbf{a}_2(\mathbf{X}_M), \dots, \mathbf{a}_P(\mathbf{X}_M)] \in \mathbb{R}^{M \times P}$  is the mixing matrix whose columns are the spatial gains of the sources at the sensor positions. For simplicity, it will be denoted as  $\mathbf{A}_M$  hereafter. Assuming that, using an unmixing matrix  $\mathbf{B} \in \mathbb{R}^{P \times M}$ , the sources are linearly estimated as  $\hat{\mathbf{s}}(t) = \mathbf{B}\mathbf{y}(\mathbf{X}_M, t)$ , the MSE,  $\mathbb{E}\{\|\mathbf{s}(t) - \hat{\mathbf{s}}(t)\|_2^2\}$ , is considered as a measure for the separation quality. The optimum matrix  $\mathbf{B}^*$  that minimizes the MSE is given by [6, Eq.(4.95)]

$$\begin{aligned} \mathbf{B}^* &= \mathbb{E}\{\mathbf{s}(t)\mathbf{y}(\mathbf{X}_M, t)^T\} \left( \mathbb{E}\{\mathbf{y}(\mathbf{X}_M, t)\mathbf{y}(\mathbf{X}_M, t)^T\} \right)^{-1} \\ &= \mathbf{A}_M^T (\mathbf{A}_M \mathbf{A}_M^T + \mathbf{C}_{MM}^n)^{-1}. \end{aligned} \quad (11)$$

Using the optimum linear estimator matrix  $\mathbf{B}^*$ , the minimum MSE is

$$\begin{aligned} \mathbb{E}\{\|\mathbf{B}^*\mathbf{y}(\mathbf{X}_M, t) - \mathbf{s}(t)\|_2^2\} &= \\ &= P - \text{Trace}(\mathbf{A}_M^T (\mathbf{A}_M \mathbf{A}_M^T + \mathbf{C}_{MM}^n)^{-1} \mathbf{A}_M). \end{aligned} \quad (12)$$

Note that  $\mathbf{A}_M$  and  $\mathbf{C}_{MM}^n$  are functions of the sensor positions, so the minimum MSE is also a function of the sensor positions. The first term in (12) is constant, so the second term is chosen as a sensor placement criterion:

$$\begin{aligned} J_{\text{MSE}} &= \text{Trace}(\mathbf{A}_M^T (\mathbf{A}_M \mathbf{A}_M^T + \mathbf{C}_{MM}^n)^{-1} \mathbf{A}_M) \\ &= \sum_{l=1}^P \mathbf{a}_l(\mathbf{X}_M)^T \left( \sum_{p=1}^P \mathbf{a}_p(\mathbf{X}_M)\mathbf{a}_p(\mathbf{X}_M)^T + \mathbf{C}_{MM}^n \right)^{-1} \mathbf{a}_l(\mathbf{X}_M). \end{aligned} \quad (13)$$

The main distinction between the final formulations of the MSE (13) and SINR (8) criteria lies in the inclusion of the term  $\mathbf{a}_l(\mathbf{X}_M)\mathbf{a}_l(\mathbf{X}_M)^T$  in summation in the inverse matrix of the MSE criterion, whereas it is absent in the SINR criterion. This slight difference may explain the closely comparable performance of these criteria, as will be demonstrated in the numerical simulations of Section 4.2.2. Another notable difference is that the MSE criterion aims to minimize the error of the estimated sources, whereas the SINR criterion focuses on eliminating the power of interference plus noise. Therefore, the MSE criterion is preferable when achieving the closest estimation to the real sources

is the goal, whereas the SINR criterion is preferred for noise and interference suppression.

To modify this criterion for the case where only the estimation of a subset of the sources (say  $\mathcal{L} \subset \{1, 2, \dots, P\}$ ) is desired, the MSE can be defined as  $\mathbb{E} \{ \|\mathbf{s}_{\mathcal{L}}(t) - \hat{\mathbf{s}}_{\mathcal{L}}(t)\|_2^2 \}$ , where  $\mathbf{s}_{\mathcal{L}}(t)$  is the vector of the sources included in  $\mathcal{L}$  at time  $t$ , and  $\hat{\mathbf{s}}_{\mathcal{L}}(t) = \mathbf{B}_{\mathcal{L}} \mathbf{y}(\mathbf{X}_M, t)$  with  $\mathbf{B}_{\mathcal{L}} \in \mathbb{R}^{|\mathcal{L}| \times M}$ . In the same way described above, using the optimum matrix of  $\mathbf{B}_{\mathcal{L}}^*$ , the MSE criterion for estimating the source subset is given by

$$J_{\text{MSE}}(\mathcal{L}) = \sum_{l \in \mathcal{L}} \mathbf{a}_l(\mathbf{X}_M)^T (\mathbf{A}_M \mathbf{A}_M^T + \mathbf{C}_{MM}^n)^{-1} \mathbf{a}_l(\mathbf{X}_M). \quad (14)$$

### 2.3. Statistical Modelling of the Spatial Gains

The criteria obtained in the previous section are functions of the spatial gains. Therefore, optimizing them to choose the best sensor locations requires information on how the spatial gains vary across the optimization space. Assuming that the spatial gains are perfectly known as a function of the sensor position is not realistic in practical applications. In real-world experiments, the spatial gains are measured at a finite set of positions, inevitably including some measurement error. Moreover, in certain applications, like EEG experiments, measurements may be conducted on a specific experimental case and then generalized to other cases. Therefore, along with information about the spatial distribution of the spatial gains, there exists uncertainty about them. To model both the prior knowledge and the uncertainty, a statistical model to represent the spatial gains is employed. The spatial gain of each source is independently modeled by a Gaussian Process (GP) as [22]

$$\hat{a}_p(\cdot) \sim \mathcal{GP}(m^{a_p}(\cdot), C^{a_p}(\cdot, \cdot)), \quad (15)$$

which implies that the  $p$ -th real spatial gain is considered as a sample function of the spatial stochastic process  $\hat{a}_p(\cdot)$ . More precisely, the above notation states that for any two points in space,  $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ , the random variables  $\hat{a}_p(\mathbf{x})$  and  $\hat{a}_p(\mathbf{x}')$  are jointly normal with covariance specified by  $C^{a_p}(\mathbf{x}, \mathbf{x}')$  and means given by  $m^{a_p}(\mathbf{x})$  and  $m^{a_p}(\mathbf{x}')$ . From the definition of Gaussian processes, for any finite number of points in space, the corresponding spatial gains follow multivariate normal distributions, where the mean vector and covariance matrix are determined by the mean and covariance functions of the GP model.

The mean function  $m^{a_p}(\cdot)$  contains the prior knowledge derived from initial measurements. In this paper, the covariance function is assumed to be in a squared exponential form, although it could be replaced with any other suitable form. The squared exponential covariance function is given by

$$C^{a_p}(\mathbf{x}, \mathbf{x}') = \sigma_p^2 \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / (2\rho_p^2)). \quad (16)$$

This covariance function is isotropic, meaning that the covariance between  $\hat{a}_p(\mathbf{x})$  and  $\hat{a}_p(\mathbf{x}')$  only depends on the distance between the points,  $\|\mathbf{x} - \mathbf{x}'\|$ . The parameter  $\sigma_p^2$  specifies the variance of  $\hat{a}_p(\mathbf{x})$  across the entire space and controls

the level of the uncertainty; a larger  $\sigma_p^2$  leads to a higher level of uncertainty. The covariance between two points decreases exponentially with increasing squared distance. The parameter  $\rho_p$  governs this rate of reduction. A smaller  $\rho_p$  leads to faster loss of correlation as distance grows, resulting in less smooth sample functions. Thus,  $\rho_p$  is termed the smoothness parameter. Determining the parameters of the GP requires some initial measurement of the spatial gains. In Section 3.2.2 an effective way of estimating the spatial gains using the measured signals of the sensors and a BSS technique will be presented.

Note that actual spatial gains can be deterministic. The rationale behind employing this stochastic model is to address the uncertainty arising from various factors, such as potential errors in the measurement of spatial gains and the finite number of sampled points in space where the spatial gains are initially measured. Having in mind the central limit theorem, it seems reasonable to use a Gaussian distribution to model uncertainties in the spatial gains. Moreover, GP model is flexible to produce various shapes for the sample functions by controlling its mean and covariance. This model also offers simplicity in its manipulation. For instance, as it will be demonstrated in Section 3.2.1, when additional estimations from spatial gains at a set of points are incorporated, updating the model and deriving the posterior distribution across the entire space is a straightforward process.

Through the statistical modeling of spatial gains, the spatial gains corresponding to sensor positions become random variables. Consequently, the criteria acquired in the previous section also transform into random variables. As a result, their expected values are employed as the objective functions for optimization with respect to the sensor locations. Specifically, the SINR criterion is expressed as

$$\hat{J}_{\text{SINR}}(\mathbf{X}_M) = \mathbb{E} \left\{ \sum_{l=1}^P \text{SINR}_l(\mathbf{f}_l^*; \mathbf{X}_M) \right\}, \quad (17)$$

while the MSE criterion takes the form

$$\hat{J}_{\text{MSE}}(\mathbf{X}_M) = \mathbb{E} \left\{ \text{Trace} \left( \mathbf{A}_M^T (\mathbf{A}_M \mathbf{A}_M^T + \mathbf{C}_{MM}^n)^{-1} \mathbf{A}_M \right) \right\}. \quad (18)$$

The expectation is similarly applied in the cases where a subset of the sources is to be estimated (9) and (14).

### 3. Optimization problem

To identify the optimal sensor locations, the criterion (based on SINR or MSE) needs to be maximized with respect to the sensor positions across the available space. When placing  $M$  sensors within a  $D$ -dimensional space, there are  $M \times D$  optimization variables, which increase with the number of sensors. The criterion includes local minima, necessitating a global optimization approach. Moreover, evaluating the criterion at specific sensor positions entails

calculating the expected value, introducing additional computational complexity. These factors collectively contribute to the high complexity of the optimization problem, making it computationally infeasible to achieve an optimal solution for practical applications.

In this section, approaches are discussed that render the optimization problem computationally solvable, albeit without the guarantee of achieving the globally optimum solution. The performances of these approaches will be numerically examined in Section 4. The first approach decomposes the optimization problem into sub-problems, in each of which the position of a single sensor is optimized. Two different methods are employed to solve these sub-problems. The second approach considers a scenario in which the sensors are placed step by step, and in each step, already placed sensors are utilized to gain new information about the spatial gains and update the stochastic model. Additionally, a method for estimating the spatial gains using sensor measurements and obtaining the prior distributions using these estimations is presented. In the third approach, the sensor positions are updated simultaneously using the gradient of the criterion to converge to a local optimum near the initial positions.

### 3.1. Greedy method

In the step by step placement approach, during each step, only the position of  $N$  sensors is optimized, while previously placed sensors are assumed to be fixed. The process begins with zero sensors placed. The optimization problem that needs to be solved in each step has  $N \times D$  variables. Therefore,  $N$  can be chosen to be low enough to ensure the sub-problem is computationally solvable. Two methods for solving the global optimization sub-problem are discussed: grid search and a gradient based global optimization method. Here, the focus will be on the case where  $N = 1$ , but it can be readily generalized to  $N > 1$ .

#### 3.1.1. Grid Search

Let  $\mathbf{X}_{K-1} \triangleq [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{K-1}]$  represent the positions of the placed sensors. In the  $K$ -th step, the position of the new sensor,  $\mathbf{x}_K$ , is being optimized. The optimization problem in this step is given by

$$\mathbf{x}_K^* = \underset{\mathbf{x}_K \in \mathcal{X}}{\operatorname{argmax}} \hat{J}(\mathbf{X}_{K-1}, \mathbf{x}_K), \quad (19)$$

where the objective function can be based on either the SINR or MSE criteria. Here,  $\mathcal{X}$  denotes the available space to place the  $K$ -th sensor. A basic approach to solve this optimization problem is to select a grid of points in the available space, denoted as  $\mathbf{X}_T = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T]$ , and evaluate the objective function at each point to find the maximum value.

As described in Section 2.3, evaluating the objective function at each point on the grid requires obtaining an expected value. However, obtaining a closed form expression for the expectation is not straightforward. Therefore, averaging over Monte Carlo realizations is employed to approximate the expected value. The spatial gain samples are obtained from their distributions and used to generate samples of the function for which the expected value is required. Subsequently, the average of these samples is used as an estimate for the expected value.

### 3.1.2. Gradient based global search

The objective function in the sub-problem (19) contains various local optima, and relying solely on a single gradient-based optimization method may lead to a local optimum. To leverage the benefits of gradient based convergence for finding the global optimum solution, a strategy is to initiate local optimizations from multiple starting points, finding the local optima near each starting point, and subsequently selecting the optimal one among them. For this purpose, starting points are chosen on a grid, and the gradient of the objective function is used for local optimization. The grid size should be sufficiently large to ensure that no local optimum is missed. However, the number of required starting points in this method to achieve the same performance as the pure grid search is significantly lower than the number of points needed to evaluate the function in the grid search approach. This will be demonstrated by the numerical results in Section 4.3.

Assume that in the  $K$ -th step of the greedy method, the positions of the placed sensors  $\mathbf{X}_{K-1}$  are fixed, and the position of the new sensor in a  $D$ -dimensional space is denoted as  $\mathbf{x}_K = (x_K^1, x_K^2, \dots, x_K^D)^T$ . The gradient of the objective function with respect to the position of the new sensor, denoted as  $\nabla_{\mathbf{x}_K} \hat{J}(\mathbf{X}_{K-1}, \mathbf{x}_K) \in \mathbb{R}^D$ , should be computed for being used in local optimization. Let us define  $j(\mathbf{X}_{K-1}, \mathbf{x}_K)$  as the spatial stochastic process whose expected value at the sensor positions represents the criterion:  $\hat{J}(\mathbf{X}_{K-1}, \mathbf{x}_K) = \mathbb{E}\{j(\mathbf{X}_{K-1}, \mathbf{x}_K)\}$ . The interchange of the expectation and gradient operations yields  $\nabla_{\mathbf{x}_K} \hat{J}(\mathbf{X}_{K-1}, \mathbf{x}_K) = \mathbb{E}\{\nabla_{\mathbf{x}_K} j(\mathbf{X}_{K-1}, \mathbf{x}_K)\}$ . The derivative of  $j(\mathbf{X}_{K-1}, \mathbf{x}_K)$  with respect to  $x_K^d$  itself constitutes a spatial stochastic process, for  $d = 1, 2, \dots, D$ . In Appendix A, the methodology for generating samples of these stochastic processes for the SINR criterion is outlined, and a similar approach can be applied to the MSE criterion.

By averaging over the samples of  $\nabla_{\mathbf{x}_K} j(\mathbf{X}_{K-1}, \mathbf{x}_K)$ , an estimate for its expected value is obtained, which represents the gradient of the criterion. This gradient is used in a gradient ascent (GA) method for local optimization starting from the initial points. Assuming  $\mathbf{x}_K^{(0)}$  is the initial point in the GA algorithm, the update rule in the  $t$ -th step of the algorithm is given by

$$\mathbf{x}_K^{(t)} = \mathbf{x}_K^{(t-1)} + \mu_{GA}^{(t)} \mathbb{E}\left\{\nabla_{\mathbf{x}_K} j(\mathbf{X}_{K-1}, \mathbf{x}_K^{(t-1)})\right\}, \quad (20)$$

where  $\mu_{GA}^{(t)}$  denotes the step size of the GA algorithm. If, at any step,  $\mathbf{x}_K^{(t)}$  lies outside the available sensor space, it is projected back into the available space.

The fact that the gradient is the expectation of a stochastic process allows us to employ a stochastic gradient ascent (SGA) algorithm. In each step of SGA, a single sample from the stochastic process, where the gradient represents its expected value, is used instead of the full estimation of the gradient. This approach may not lead to an increase in the objective function at every step, but overall, it can converge to a stationary point. In the literature, the convergence of the SGA algorithm to a stationary point is shown under some strong assumptions on the objective function [23]. In this paper, as with many machine

learning models employing SGA as the optimization algorithm, the convergence will be verified numerically. The update rule for the variable sensor position  $\mathbf{x}_K$  in the  $t$ -th step of SGA is then

$$\mathbf{x}_K^{(t)} = \mathbf{x}_K^{(t-1)} + \mu_{SGA}^{(t)} \nabla_{\mathbf{x}_K} \hat{j}(\mathbf{X}_{K-1}, \mathbf{x}_K^{(t-1)}), \quad (21)$$

where  $\nabla_{\mathbf{x}_K} \hat{j}$  indicates one sample of the stochastic process.

### 3.2. Sequential approach

Recall the greedy method, wherein sensors are placed step by step. In certain applications, previously placed sensors may be utilized to acquire new information about the spatial gains, which can then improve the placement of the remaining sensors. This is called sequential approach in [17]. Here, first the updating rule of the stochastic model based on the estimation provided by the placed sensors is presented. Subsequently, a BSS estimation method for the spatial gains using sensor data is discussed.

#### 3.2.1. Stochastic model update

Let us assume that the placed sensors at the positions  $\mathbf{X}_{K-1} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{K-1}]$  have provided an estimate of the spatial gains at  $\mathbf{X}_{K-1}$ , denoted by  $\mathbf{z}_p(\mathbf{X}_{K-1})$  for the  $p$ -th source. The estimation error  $\mathbf{v}_p(\mathbf{X}_{K-1})$  is assumed to be Gaussian and independent of the spatial gains. For  $p = 1, 2, \dots, P$ , we have

$$\mathbf{z}_p(\mathbf{X}_{K-1}) = \hat{\mathbf{a}}_p(\mathbf{X}_{K-1}) + \mathbf{v}_p(\mathbf{X}_{K-1}). \quad (22)$$

Assuming that  $\mathbf{x}_K$  is the new sensor position and  $\mathbf{X}_K \triangleq [\mathbf{X}_{K-1}, \mathbf{x}_K]$ , because of the joint normal distribution of the random vector  $\begin{bmatrix} \mathbf{z}_p(\mathbf{X}_{K-1}) \\ \hat{\mathbf{a}}_p(\mathbf{X}_K) \end{bmatrix}$ , the posterior distribution of  $\hat{\mathbf{a}}_p(\mathbf{X}_K)$  is given by

$$\hat{\mathbf{a}}_p(\mathbf{X}_K) | \mathbf{z}_p(\mathbf{X}_{K-1}) \sim \mathcal{N}(\mathbf{m}_{\mathbf{a}|\mathbf{z}}^p, \mathbf{C}_{\mathbf{a}|\mathbf{z}}^p), \quad (23)$$

with

$$\begin{aligned} \mathbf{m}_{\mathbf{a}|\mathbf{z}}^p &= \mathbb{E} \{ \hat{\mathbf{a}}_p(\mathbf{X}_K) | \mathbf{z}_p(\mathbf{X}_{K-1}) \} = \mathbf{m}_K^{a_p} + \\ &\mathbf{C}_{KK}^{a_p} (\mathbf{C}_{K-1K-1}^{a_p} + \mathbf{C}_{K-1K-1}^v)^{-1} (\mathbf{z}_p(\mathbf{X}_{K-1}) - \mathbf{m}_{K-1}^{a_p}) \end{aligned} \quad (24)$$

and

$$\begin{aligned} \mathbf{C}_{\mathbf{a}|\mathbf{z}}^p &= \mathbb{E} \left\{ \|\hat{\mathbf{a}}_p(\mathbf{X}_K) - \mathbf{m}_{\mathbf{a}|\mathbf{z}}^p\|_2^2 | \mathbf{z}_p(\mathbf{X}_{K-1}) \right\} = \\ &\mathbf{C}_{KK}^{a_p} - \mathbf{C}_{KK}^{a_p} (\mathbf{C}_{K-1K-1}^{a_p} + \mathbf{C}_{K-1K-1}^v)^{-1} (\mathbf{C}_{KK-1}^{a_p})^T, \end{aligned} \quad (25)$$

where  $\mathbf{m}_K^{a_p}$  represents the mean of  $\hat{\mathbf{a}}_p(\mathbf{X}_K)$ ,  $\mathbf{C}_{K-1K-1}^{a_p}$  represents the covariance matrix of  $\hat{\mathbf{a}}_p(\mathbf{X}_{K-1})$ , and  $\mathbf{C}_{KK-1}^{a_p}$  represents the cross-covariance between  $\hat{\mathbf{a}}_p(\mathbf{X}_K)$  and  $\hat{\mathbf{a}}_p(\mathbf{X}_{K-1})$ , all of which are obtained from the GP model. The matrix  $\mathbf{C}_{K-1K-1}^v$  denotes the covariance matrix of the estimation error  $\mathbf{v}_p(\mathbf{X}_{K-1})$ .

Having obtained the posterior distribution of the spatial gains for the placement of the new sensor, the criterion will be computed by averaging over Monte Carlo realizations based on this distribution, and the optimization problem at each step of the placement is

$$\mathbf{x}_K^* = \operatorname{argmax}_{\mathbf{x}_K \in \mathcal{X}} \hat{J}(\mathbf{X}_K | \mathbf{z}_1(\mathbf{X}_{K-1}), \dots, \mathbf{z}_P(\mathbf{X}_{K-1})). \quad (26)$$

### 3.2.2. BSS estimation method

One approach to estimate spatial gains involves utilizing sensor measurements and applying a BSS technique to separate the sources, followed by estimating the mixing matrix. In some applications, such as EEG experiments, this method would be the only practical way of the spatial gains estimation. There could be several reasons for this limitation: it is not possible to deactivate the propagation of all sources except the intended one for measuring its propagation attenuation, or it is not straightforward to model the propagation environment to estimate the gains. This estimation method can also be used to gain the prior information of the spatial gains to estimate the parameters of the GP model (i.e. the mean and the covariance functions).

Consider  $K$  placed sensors measuring  $N_s$  samples. According to (2), the measurements can be expressed as

$$\mathbf{y}(\mathbf{X}_K, t) = \mathbf{A}_K(\mathbf{X}_K)\mathbf{s}(t) + \mathbf{n}(\mathbf{X}_K, t), \quad t = 1, \dots, N_s, \quad (27)$$

where  $\mathbf{A}_K(\mathbf{X}_K) = [\mathbf{a}_1(\mathbf{X}_K), \mathbf{a}_2(\mathbf{X}_K), \dots, \mathbf{a}_P(\mathbf{X}_K)] \in \mathbb{R}^{K \times P}$  represents the mixing matrix. This estimation is carried out when the number of placed sensors is greater than or equal to the number of sources ( $K \geq P$ ).

The estimation process comprises two stages. In the first stage, the objective is to find a transformation matrix to reduce measurement dimensionality while minimizing noise influence. This is achieved through Principal Component Analysis (PCA), in which the rows of the transformation matrix are chosen to be the eigenvectors of the covariance matrix of the measurements [24].

In the second stage, a BSS method for noisy measurements, like fastICA [24], separates the sources and estimates the mixing matrix blindly, providing an estimation of the spatial gains for all the sources at the sensor positions. This estimation is modeled as  $\mathbf{z}_p(\mathbf{X}_K)$  in (22) for use in the placement procedure. The estimation error  $\mathbf{v}_p(\mathbf{X}_K)$  is approximated as a Gaussian random vector with zero mean, and its variance is estimated numerically by repeating the estimation of the spatial gains for independent generation of them with the same GP parameters and obtaining the sample variance of the error.

### 3.3. Refining sensor positions from initial points

In the third approach sensor locations are simultaneously updated using the gradient of the objective function. As described in Algorithm 1, in each iteration, the positions of the sensors are updated one by one based on the gradient of the criterion with respect to the position of the corresponding sensor. It is important to note that the sensor positions will be refined, ultimately

converging to a local optimum of the criterion near the initial points. Hence, the choice of initial points greatly influences the quality of the optimization outcome.

---

**Algorithm 1**


---

**Input:** Initial sensor positions  $\mathbf{X}_M^{(0)} = [\mathbf{x}_1^{(0)}, \dots, \mathbf{x}_M^{(0)}]$ .

**Output:** Refined sensor positions.

1: **For**  $t = 1 : T$

2:   **For**  $k = 1 : M$

4:      $\mathbf{X}_k^{(t)} \leftarrow [\mathbf{x}_1^{(t)}, \mathbf{x}_2^{(t)}, \dots, \mathbf{x}_{k-1}^{(t)}, \mathbf{x}_{k+1}^{(t-1)}, \dots, \mathbf{x}_M^{(t-1)}]$

3:      $\mathbf{x}_k^{(t)} \leftarrow \mathbf{x}_k^{(t-1)} + \mu_{GA}^{(t)} \mathbb{E}\{\nabla_{\mathbf{x}_k} j(\mathbf{X}_k^{(t)}, \mathbf{x}_k^{(t-1)})\}$

5: **Return**  $\mathbf{X}_M^{(T)} = [\mathbf{x}_1^{(T)}, \dots, \mathbf{x}_M^{(T)}]$

---

#### 4. Numerical Experiments

In this section, numerical results will be presented to assess the performance of the proposed optimal sensor placement methods. The general experimental setup is detailed in Section 4.1. Section 4.2 will compare the proposed criterion with other sensor placement methods, including SNR based criteria [16, 17] and classical approaches such as entropy [25] and mutual information (MI) [26, 27] criteria. Section 4.3 will evaluate and compare different optimization approaches. Finally, in Section 4.4, the impact of the model parameters on the output performance will be examined. In addition, Section 4.5 illustrate the interest and the behavior of the proposed method on actual brain-computer interface (BCI) data. The simulations in this section are done in Matlab-R2017b on a Windows 7 operating system, with a 2.2 GHz Core i7 processor and 8 GB DDR3 memory.

##### 4.1. Experiment setup

The sources are to be placed in a 2-dimensional space within a square defined by  $\mathbf{x} \in [0, 1]^2$ . The covariance matrix of the GP model for the spatial gains of the sources is assumed to have the form given by (16). Both the variance parameter  $\sigma_u$  and the smoothness parameter  $\rho_u$  for all sources are equal. They are denoted with a subscript  $u$  to signify their association with the uncertainty regarding the spatial gains.

The mean function for each source, which provides the prior knowledge, is generated as a sample function from a GP model with a zero mean function and the covariance function parameters  $\sigma_a$  and  $\rho_a$  for all sources. The smoothness parameter  $\rho_a$  is selected to be consistent with the smoothness of the covariance function for the spatial gains,  $\rho_u$ . Note that the mean function can take other forms; however, for simplicity, it is generated in this way. The noise covariance function is also assumed to have the squared exponential form:  $C^n(\mathbf{x}, \mathbf{x}') = \sigma_n^2 \exp(-\|\mathbf{x} - \mathbf{x}'\|^2 / (2\rho_n^2))$ . In practical experiments, it is more likely that the noise is less correlated than the spatial gains. Therefore, in all simulations, the

ratio of the smoothness parameter of the noise to that of the spatial gains is set as  $\rho_n/\rho_a = 0.2$ . The impact of this ratio will be examined in Section 4.4. The sigma parameter for both the spatial gains and the noise is set to  $\sigma_a = \sigma_n = 1$  (i.e., input SNR is 0dB).

In all the figures, to obtain the mean and standard deviation of the plotted values, 100 Monte Carlo simulations are performed, in each of which a new generation of the spatial gains from the GP model is used. Note that when calculating the output SINR and MSE at the placed sensor positions, the linear estimator vector of  $\mathbf{f}_p$  and the unmixing matrix of  $\mathbf{B}$  (described in Section 2.2) are obtained from the available estimation of the spatial gains at the sensor positions, not their real values. In the sequential approach, the BSS estimation or the perfect estimation (i.e., true spatial gains) is employed, while in the greedy method, the mean of the GP is employed as the available estimation. Moreover, in any experiment that the grid search is used, 20 Monte Carlo realizations are employed to obtain the expected value in the evaluation of the objective function.

## 4.2. Comparing with other sensor placement methods

### 4.2.1. Single source extraction

The recent works of [16] and [17] have focused on optimal and robust sensor placement for extracting a single source from noisy measurements. In [16], the placement criterion is based on the expectation of the SNR of the extracted signal. In [17], the criterion is the probability that the SNR exceeds a certain threshold, and it is showed that the latter criterion is more robust against uncertainties in the spatial gains compared with the former one. To compare the proposed criteria with these works, an experiment is conducted in which there are 5 sources with additive noise. The objective is to extract one source from the noisy mixed measurements. Therefore, the proposed SINR and MSE criteria are employed, assuming that the subset of the desired sources includes only a single source.

Additionally, two other classical sensor placement criteria are incorporated: entropy and MI [25, 26, 27]. These criteria assume that each point in space corresponds to certain stochastic variables (spatial gains in our problem), and aim to sample among them optimally. The entropy criterion chooses the most informative points by maximizing their entropy, and the MI criterion ensures that the selected points best represent the rest by maximizing the mutual information between them and the remaining points [22].

In this experiment, the sigma parameter of the uncertainty is  $\sigma_u = 0.3$ , and the smoothness parameters are set to  $\rho_a = \rho_u = 0.05$ . Similar to the methodology used in [17], a sequential approach is employed, assuming perfect estimation of the spatial gains at each step (i.e., no error is introduced). The optimization is conducted using a grid search with a grid size of  $40 \times 40$ . Fig. 2 shows the output SINR of the extracted source with respect to the number of the sensors placed using these 5 criteria. It can be observed that the proposed SINR and MSE criteria have the same performance in this experiment, achieving the

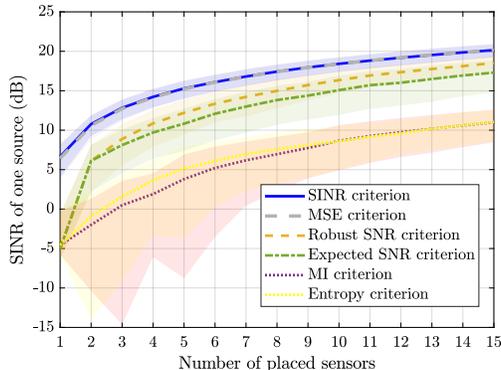


Figure 2: Performance comparison of the proposed criteria with previous source extraction criteria, robust and expected SNR, and classical criteria, MI and entropy. The SINR of the extracted source is plotted versus the number of the placed sensors, and the sequential approach with perfect estimation of the spatial gains in each step is used. The curves (resp. shadows) represent the average value (resp. standard deviation around the mean) obtained from Monte Carlo simulations.

highest SINR for each number of placed sensors. With 15 placed sensors, the proposed criteria yields a mean SINR of 20.1 dB. In contrast, the robust SNR criterion yields 18.5 dB, the expected SNR criterion yields 17.3 dB, the MI criterion yields 11.0 dB, and the entropy criterion yields 11.1 dB. Looking at it from another perspective, to attain a mean SINR of 15 dB, only 5 sensors are needed using the proposed criteria, whereas the robust and expected SNR criteria require 8 and 10 sensors, respectively. The SNR-based criteria are constrained to extract a single source. In the presence of other sources, they exhibit inferior performance compared with the SINR criteria, as they do not leverage information about the remaining spatial gains. Both the MI and entropy criteria demonstrate similar, significantly poorer performance than the other criteria in terms of output SINR. This may be attributed to their neglect of noise effects, as they primarily focus on acquiring samples that yield most information about spatial gains.

#### 4.2.2. Source separation

In this experiment, two proposed criteria, SINR (Eq. (17)) and MSE (Eq. (18)), are compared with the classical criteria, MI and entropy. The model parameters are set to  $\sigma_u = 0.3$ ,  $\rho_a = \rho_u = 0.05$ . For optimization, a greedy method with a grid search of size  $40 \times 40$  is employed, assuming the existence of  $P = 3$  sources. Figs. 3a and 3b depict the average of the SINRs of the separated sources and their output MSE, respectively, with respect to the number of placed sensors. It can be observed that the MSE and SINR criteria exhibit nearly identical performances, both surpassing the classical criteria. After placing 15 sensors, the SINR and MSE criteria yield the mean averaged SINRs of 16.8 dB

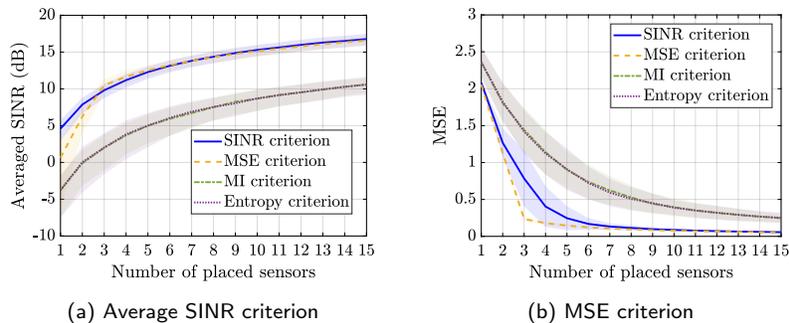


Figure 3: The performances of the SINR and MSE criteria in sensor placement for source separation are compared with the classical sensor placement criteria. Two measures are used to assess the quality of source separation: (a) Average of the SINRs of the separated sources. (b) MSE of the source separation. The curves (resp. shadows) represent the average value (resp. standard deviation around the mean) obtained from Monte Carlo simulations.

and 16.5 dB, respectively, whereas the MI and entropy criteria both yield the mean averaged SINRs of 10.6 dB. The mean output MSE for the SINR and MSE criteria after placing 15 sensors reaches 0.057 and 0.054, respectively, while for the MI and entropy criteria, it reaches 0.246 and 0.251, respectively. It is evident that the SINR criterion yields slightly better results when the performance measure is the averaged SINR (Fig. 3a). Conversely, the reverse holds true when the measure is MSE (Fig. 3b). This outcome aligns with expectations, as each criterion is designed to optimize its respective performance measure. Furthermore, the closely matched performances of these two criteria suggests a high degree of similarity between them. Given this similarity, and to avoid repetition, we will exclusively employ one of the two criteria in the remaining experiments. We choose the SINR criterion because obtaining its gradient is simpler than for the MSE criterion.

The sensor positions chosen by the four criteria using the greedy method, along with the real spatial gains of the three sources, are shown in Fig. 4. Notably, both the SINR and MSE criteria have led to the placement of some sensors at the same locations, while some others are placed in close proximity. However, distinct placements are also observed. On the other hand, the entropy criterion resulted in sensors being positioned far apart, with extending some of them to the borders. This characteristic can be considered a drawback of the entropy criterion, as sensors typically gather information from their surroundings, making a sensor at the boundary less efficient in utilizing sensed data [27]. Note that the sensor positions determined by the MI and entropy criteria remain fixed in each generation of the spatial gains from the GP model. This is because these two criteria solely depend on the covariance function of the GP, without utilizing the prior knowledge provided by the mean function. This limitation is a significant drawback of these criteria.

In Fig. 5, the averaged SINR of the sources is plotted after placing 15 sensors,

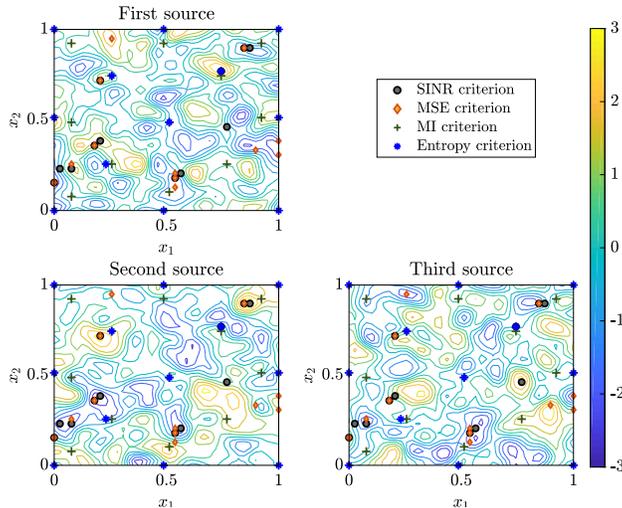


Figure 4: Contour plots show the real spatial gain values for each of the three sources. The positions of 12 sensors, determined by four criteria, are indicated on the plots.

as a function of the number of existing sources. It is evident that an increase in the number of sources leads to a decrease in the averaged SINR. This decline can be attributed to the fact that in the estimation of each source, the remaining sources act as interference. As the number of sources increases, the power of this interference correspondingly rises, leading to a reduction in the averaged SINR.

#### 4.3. Evaluating different sensor placement and optimization approaches

##### 4.3.1. Gradient based optimization: gradient ascent vs. stochastic gradient ascent

Two local optimization methods, gradient ascent (GA) and stochastic GA (SGA), were presented in Section 3.1.2 to be used in the steps of the gradient based global optimization. Fig. 6 compares the convergence of these methods from the starting points chosen on an  $8 \times 8$  grid. The contour plot displays the SINR objective function for placing the second sensor ( $K = 2$ ) across a 2-D space. The variable positions  $\mathbf{x}_K^{(t)}$  in all iterations are represented by black dots. The model parameters are set to  $\sigma_u = 0.3$ ,  $\rho_a = \rho_u = 0.05$  and  $P = 3$ . The step size of SGA is chosen to be smaller than that of GA due to the inaccuracy of the steps in SGA. This implies the need for more steps in SGA compared with GD to converge to a local maximum. On the other hand, computing the update direction in GA requires a higher computational cost than in SGA, as it involves averaging over several samples of the gradient instead of using just one.

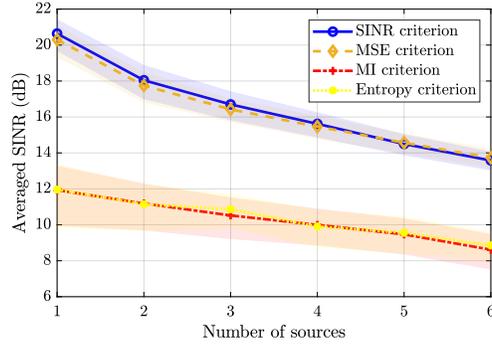


Figure 5: Influence of the number of sources on the averaged SINR for 15 sensors. The curves (resp. shadows) represent the average value (resp. standard deviation around the mean) obtained from Monte Carlo simulations.

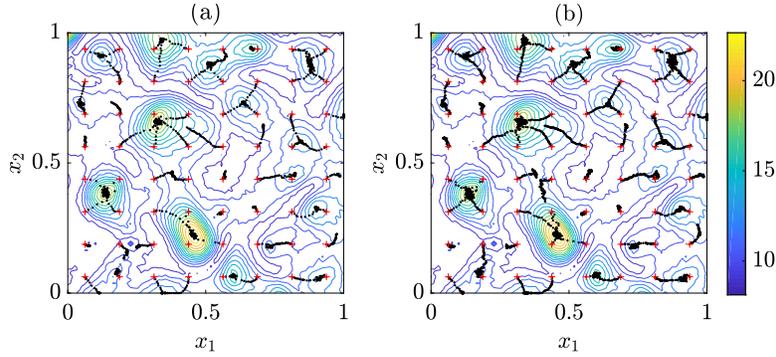


Figure 6: Comparison of gradient ascent and stochastic gradient ascent algorithm for optimal sensor placement. The contour plot illustrates the SINR objective function for  $K = 2$ . The initial points of the local optimization algorithm ( $\mathbf{x}_K^{(0)}$ ) are denoted by red plus signs, while the updates of the variable position in different iterations ( $\mathbf{x}_K^{(t)}$ ) are represented by black dots. Two algorithms for local optimization: (a) GD with a step size of  $\mu_{GA}^{(t)} = 0.0002$ , executed over 30 iterations, with averaging performed over 5 Monte Carlo simulations in each iteration. (b) SGD with a step size of  $\mu_{SGA}^{(t)} = 0.00008$ , executed over 80 iterations.

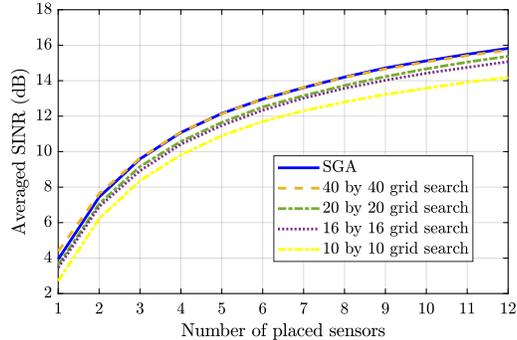


Figure 7: Comparing the accuracy obtained by the gradient based global optimization using SGA with the grid search using different grid sizes. The gradient based global optimization method achieves the accuracy similar to that of a 40 by 40 grid search.

#### 4.3.2. Grid search vs. gradient based global optimization

In this experiment, two global optimization methods introduced in Section 3.1 are compared for solving the optimization problem at each step of the greedy method. In the grid search, 20 Monte Carlo realizations are used to obtain the expectation in the evaluation of the SINR criterion at each point on the grid. For the local optimization step of the gradient based global optimization method, the SGA algorithm with 60 iterations is employed. During the initial 45 iterations, the step size is set as  $\mu_{SGA}^{(t)} = 2 \times 10^{-4}$  for a rapid convergence, and in the final 15 iterations, it is adjusted to  $\mu_{SGA}^{(t)} = 5 \times 10^{-5}$  for a more precise convergence. The starting points for the gradient based global optimization method are selected on an  $8 \times 8$  grid. These method parameters are determined empirically through multiple simulations to optimize the performance. The model parameters are set as follows:  $\rho_a = \rho_u = 0.05$ ,  $\sigma_u = 0.3$ , and  $P = 3$ . In Fig. 7, a comparison is made between the resulting averaged SINR values obtained using the SGA based global optimization and the grid search with different grid sizes. Additionally, Table 1 provides a comparison of the required time for placing 15 sensors using these algorithms.

It can be observed that the accuracy of the SGA is comparable to that of a  $40 \times 40$  grid search, while its runtime is reduced by a factor of 8, making it comparable to that of a  $15 \times 15$  grid search. This indicates that SGA provides significant computational efficiency compared with the grid search. Note that the number of grid points in the grid search grows exponentially with the dimension of space or the number of placed sensors at each step of the greedy method. Therefore, for higher space dimensions or a greater number of placed sensors, the efficiency of the gradient based optimization will become even more pronounced.

Table 1: Required Time for Placing 15 Sensors

Method	SGA	Grid search			
		$10 \times 10$	$16 \times 16$	$20 \times 20$	$40 \times 40$
Time(s)	12.4	5.2	15.7	24.8	103.2

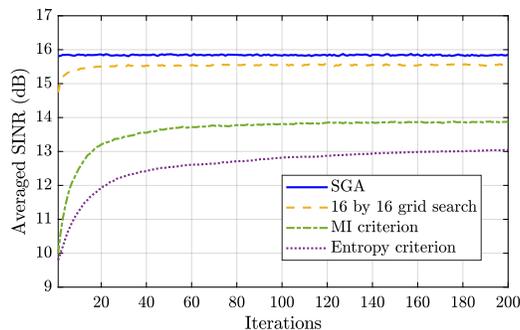


Figure 8: Effect of simultaneously updating the sensor positions from 4 different initial positions onto the SINR. Improvement of the resulting averaged SINR is plotted against the number of iterations of the algorithm.

#### 4.3.3. Refining sensor positions from initial points

In this experiment, the optimization approach outlined in Section 3.3 is evaluated, wherein sensor positions are updated step by step from an initial point using the gradient of the objective function. The model parameters utilized are  $\rho_a = \rho_u = 0.05$ ,  $\sigma_u = 0.3$ , and  $P = 3$ .

Fig. 8 compares the effect of different starting sensor positions on the mean of the resulting averaged SINR during the iterations of Algorithm 1, which aims at adjusting the sensor positions. When the initial positions are derived from the output of the greedy method with SGA, it is evident that the algorithm does not lead to an improvement in sensor positions. This indicates that these initial positions are already situated at a local optimum of the objective function with respect to all sensor positions. However, when the initial positions are derived from the output of the greedy method employing a  $16 \times 16$  grid search, the algorithm refines the sensor positions and converges to a local optimum close to the initial points, but cannot reach the performance achieved by the SGA. Furthermore, it is observed that the algorithm shows greater enhancement in the initial points chosen based on the MI criterion compared with those chosen using the entropy criterion. This observation can be attributed to the fact that some initial points from the entropy criterion are positioned at the borders of the available space. As a result, the local optima near them are less optimal than those identified by the MI criterion. This experiment highlights the effective performance of the SGA in our problem.

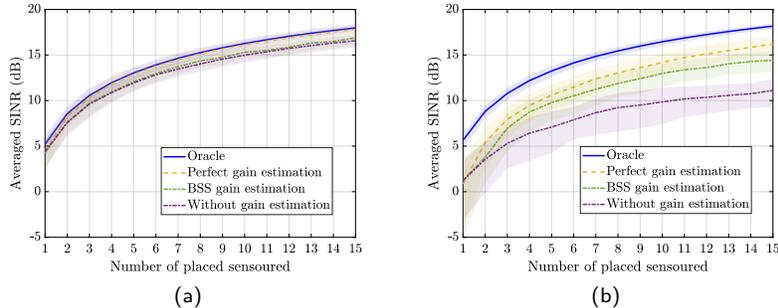


Figure 9: Comparing the sequential approach, where perfect or BSS estimation of the spatial gains is performed in each step, with the greedy method, where the knowledge about the spatial gains remains constant across steps. Two cases are considered: (a) Low level of uncertainty with  $\sigma_u = 0.3$ , and (b) High level of uncertainty with  $\sigma_u = 0.8$ .

#### 4.3.4. Greedy method vs. sequential approach

As described in Section 3.2, the distinction between the greedy method and the sequential approach lies in the availability of an estimation of the spatial gains at the placed sensor positions at each step. This estimation is utilized to update the stochastic model of the spatial gains in the sequential approach, whereas it is not available in the greedy method. In this experiment, two types of estimations are considered for the sequential approach: the BSS method outlined in Section 3.2.2 and perfect estimation. In the case of perfect estimation, there are no errors in the estimation of the spatial gains. This case is included to evaluate the impact of BSS estimation errors on the performance of the sensor placement.

Additionally, an oracle case is considered, which assumes no uncertainty about the spatial gains and treats them as deterministic. This can serve as an upper bound for the resulting averaged SINR of the proposed method, and specified by the solid (blue) curve in Fig. 9.

For this experiment, the smoothness parameters are set to  $\rho_a = \rho_u = 0.05$ , the number of sources is  $P = 3$ , and two levels of uncertainty are simulated. The low uncertainty level has  $\sigma_u = 0.3$ , while the high uncertainty level has  $\sigma_u = 0.8$ . As illustrated in Fig. 9a, under the low uncertainty level of the spatial gains, the BSS estimation in the sequential approach does not provide a significant advantage compared with the greedy method (without gain estimation). This can be attributed to the fact that the uncertainty and the BSS estimation error fall within similar ranges. However, under the high uncertainty level of the spatial gains in Fig. 9b, the information provided by the BSS gain estimation leads to a higher averaged SINR, increasing it from 11.1 dB to 14.4 dB for 15 placed sensors. The gap between the perfect and BSS gain estimations illustrates the effect of the error in BSS estimation.

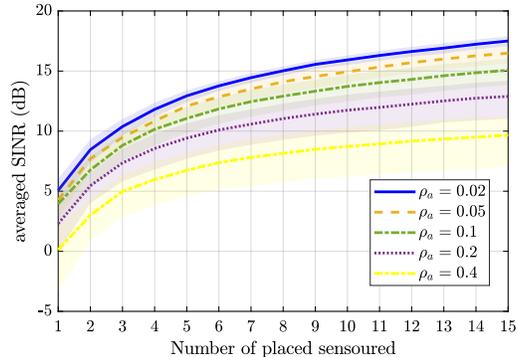


Figure 10: Influence of the spatial gain smoothness. The curves (resp. shadows) represent the average SINR value (resp. standard deviation around the mean) obtained from Monte Carlo simulations.

#### 4.4. Impact of model parameters

##### 4.4.1. Smoothness of the spatial gains

Fig. 10 illustrates the resulting averaged SINR for the sensor placement using the greedy method with different smoothness levels of the spatial gains, when the ratio of the smoothness parameter of the noise to the spatial gains is fixed to  $\rho_n/\rho_a = 0.2$ . The other model parameters are set as follows:  $\rho_u = \rho_a$ ,  $\sigma_u = 0.3$ ,  $P = 3$ . Decreasing  $\rho_a$  reduces the smoothness of the spatial gains, thereby increasing their variability within the available space, resulting in an increase in the averaged SINR. This aligns with the fundamental moral notion of diversity in BSS. Furthermore, at a high smoothness level, the output averaged SINR saturates faster as the number of placed sensors increases. It appears that the capacity of the space for accommodating sensors decreases with the increase in the smoothness of the spatial gains.

##### 4.4.2. Noise smoothness to spatial gain smoothness ratio

In this experiment, the ratio of the smoothness parameter of the noise to the smoothness parameter of the spatial gains ( $\rho_n/\rho_a$ ) is varied from 0.1 to 10, and the averaged SINR after placing 15 sensors using the SINR and MI criteria is plotted in Fig. 11. The other model parameters are set as  $\rho_u = \rho_a = 0.05$ ,  $\sigma_u = 0.3$ ,  $P = 3$ , and the sequential approach with perfect gain estimation is employed. It can be observed that as the ratio exceeds 1 and the noise becomes smoother than the spatial gains, the resulting averaged SINR of the SINR criterion begins to increase. In this scenario, the criterion selects sensor locations that are in close proximity to each other, at a distance where the noise signals exhibit high correlation, while the spatial gains have lower correlations. This enables the effective suppression of the noise signal without compromising the source signals, resulting in a high achievable output SINR.

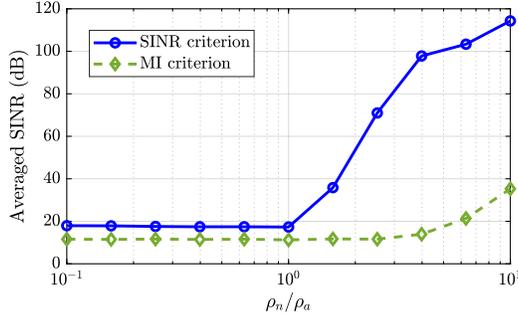


Figure 11: Influence of the ratio between the smoothness of the noise and the smoothness of the spatial gain. The curves (resp. shadows) represent the average value after placing 15 sensors (resp. standard deviation around the mean) obtained from Monte Carlo simulations.

#### 4.5. Real-data verification: P300-based BCI dataset

The proposed method was applied to actual EEG data from the BCI database by [28] to evaluate the impact of optimal sensor placement on the performance of the BCI system. The database includes recorded EEG data from 54 subjects across 2 sessions, using 62 electrodes placed on the scalp. This setup provides a diverse range of subjects with high spatial resolution. The BCI experiment based on the P300 event-related potential (ERP) [29] implemented in the database follows a typical 36-symbol Row-Column speller design. The objective is to detect P300 ERPs in the EEG signal and use them to specify the spelled symbol. To enhance the detectability of P300 ERPs, we employ the xDAWN algorithm [30]. This algorithm decomposes the EEG signals into a linear mixture of sources of interest plus additive noise and designs optimal spatial filters to estimate the sources. According to this model, the EEG signal is expressed as

$$\mathbf{X} = \mathbf{DA}'\mathbf{W}^T + \mathbf{N}', \quad (28)$$

where the columns of  $\mathbf{X} \in \mathbb{R}^{N_t \times N_s}$  represent the recorded EEG signals using  $N_s$  electrodes in  $N_t$  time samples. The columns of  $\mathbf{DA}' \in \mathbb{R}^{N_t \times I}$  specify  $I$  sources of interest, the columns of  $\mathbf{W} \in \mathbb{R}^{N_s \times I}$  denote the source gains in electrodes, and  $\mathbf{N}' \in \mathbb{R}^{N_t \times N_s}$  denotes the noise signal. The matrices  $\mathbf{DA}'$  and  $\mathbf{W}$  can be learned using the training data, as explained in [30]. The sources of interest are estimated and utilized as classification features to detect the P300 ERPs. The optimal placement method aims to predict the sensor positions that provide the best estimation of sources.

For each subject, the training data from the first session is used to obtain the model parameters, and then the optimal electrodes are selected based on the model. The number of sources is chosen as  $I = 2$ , as it has been shown that 2 sources provide sufficient features for classification in P300 spellers [30]. To estimate the parameters of the model, samples of the stochastic process of the spatial gains are required. To achieve this, a sliding window with a length equal to half of the time samples ( $N_t/2$ ) is used to select segments of the signal, and

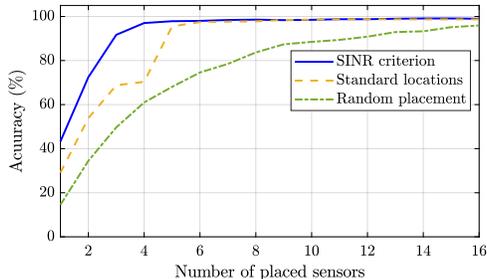


Figure 12: BCI accuracy obtained from electrodes configured using three placement approaches, averaged over 40 subjects

each selected segment is used to obtain one sample for the spatial gains using the xDAWN algorithm. For each subject, a total of 10 samples of the stochastic process are obtained at 62 electrode locations ( $\hat{\mathbf{W}}^{(1)}, \hat{\mathbf{W}}^{(2)}, \dots, \hat{\mathbf{W}}^{(10)}$ ). The columns of  $\hat{\mathbf{W}}^{(i)}$  are normalized so that the corresponding source signals have unit variances. The mean of the stochastic process is estimated by averaging these samples, and the parameters of the isotropic covariance function are obtained using the least-squares estimation of the Gaussian variogram [31]. The noise covariance matrix is estimated using the noise signal samples obtained with the xDAWN algorithm from the entire training data, as  $\mathbf{C}^n = \frac{1}{N_t} \mathbf{N}'^T \mathbf{N}'$ . With this information, the optimal electrode positions are predicted using the SINR criterion and the greedy optimization method. Note that the available electrode positions are limited to the locations of the 62 electrodes used in the database. The chosen electrode positions from the training data of the first session are used to linearly estimate the sources in the second session. Subsequently, the estimated sources serve as features for a Bayesian linear discriminant analysis (BLDA) classifier [32, 33]. The classification accuracy is compared with the accuracy obtained from random electrode placement and the standard electrode configurations commonly used in P300-based BCI systems [32, 34, 35]. The standard 4-electrode configuration includes Fz, Cz, Pz, and Oz electrodes. In the 8-electrode configuration, P3, P4, P7, and P8 electrodes are added. The 16-electrode configuration includes 8 additional electrodes: C3, C4, O1, O2, CP1, CP2, FC1, and FC2. Fig. 12 compares the classification accuracy obtained from these three electrode placement approaches. For each subject, the classifier is trained and tested using the second session data, and the reported accuracy is the average of 40 subjects. Optimal electrode placement results in higher accuracy than standard placement, especially when the number of sensors is below 6. With 4 placed electrodes, the accuracy obtained by optimal sensor placement is 97.0%, while it is 70.3% by standard placement. With 8 placed electrodes, the accuracy is 98.5% and 97.8% respectively. Both optimal and standard placements achieve higher accuracy than randomly placing the electrodes on the scalp.

## 5. Conclusion

In this paper, by assuming a linear mixing model of the sources with noisy measurements, two criteria have been proposed as measures for the separation quality to be optimized in the sensor placement process. To integrate the knowledge and uncertainty about the spatial gains, they were modeled as GPs, resulting in stochastic forms for the proposed criteria and the optimization problem. To efficiently solve the global optimization problem required in the placement process, a gradient based optimization method has been used that provides significantly greater computational efficiency compared with the previously used grid search method.

Numerical results demonstrated the superiority of the proposed criteria compared with the criteria of [16] and [17] in source extraction, as well as the classical sensor placement criteria in source separation. Additionally, we observed that the SINR and MSE criteria yield similar performances in the numerical experiments. The approach for updating sensor positions from specific initial points improves the sensor positions, although the final separation quality is highly dependent on the initial points chosen. The impact of smoothness model parameters for spatial gains and noise was also explored.

As a real-world application of the proposed method, it was verified using real data from a P300-based BCI dataset. The optimal placement method demonstrated higher accuracy compared to standard locations. This leads to the need for fewer electrodes in the BCI system to achieve a certain level of accuracy, improving the ergonomics of the BCI by reducing the setup time and facilitating the frequent use of the system.

The instantaneous mixing model of the sources has been addressed in this paper. In applications where echo signals are present in the environment, such as acoustic signals, the convolutive mixing model is relevant. A follow-up to the current work could be to study the optimal sensor placement problem with convolutive mixtures and noisy measurements.

## Appendix A.

For the SINR criterion, the stochastic process  $j(\mathbf{X}_{K-1}, \mathbf{x}_K)$  is given by

$$j(\mathbf{X}_{K-1}, \mathbf{x}_K) = \sum_{l=1}^P \hat{\mathbf{a}}_l(\mathbf{X}_K)^T \times \left( \sum_{p=1, p \neq l}^P \hat{\mathbf{a}}_p(\mathbf{X}_K) \hat{\mathbf{a}}_p(\mathbf{X}_K)^T + \mathbf{C}_{KK}^n \right)^{-1} \hat{\mathbf{a}}_l(\mathbf{X}_K), \quad (\text{A.1})$$

where  $\mathbf{X}_K = [\mathbf{X}_{K-1}, \mathbf{x}_K]$ . Among the spatial gains of the  $p$ -th source,  $\hat{\mathbf{a}}_p(\mathbf{X}_M) = [\hat{a}_p^1(\mathbf{x}_1), \hat{a}_p^2(\mathbf{x}_2), \dots, \hat{a}_p^M(\mathbf{x}_K)]$ , only  $\hat{a}_p^K(\mathbf{x}_K)$  depends on  $\mathbf{x}_K$ . By applying the chain rule, the partial derivatives of  $j(\mathbf{X}_K)$  with respect to  $x_K^d$

( $d = 1, 2, \dots, D$  for  $D$ -dimensional space) can be expressed as

$$\begin{aligned} \frac{\partial j(\mathbf{X}_K)}{\partial x_K^d} &= \frac{\partial j(\mathbf{X}_K)}{\partial \hat{a}_1^K(\mathbf{x}_K)} \cdot \frac{\partial \hat{a}_1^K(\mathbf{x}_K)}{\partial x_K^d} + \frac{\partial j(\mathbf{X}_K)}{\partial \hat{a}_2^K(\mathbf{x}_K)} \cdot \frac{\partial \hat{a}_2^K(\mathbf{x}_K)}{\partial x_K^d} + \dots \\ &+ \frac{\partial j(\mathbf{X}_K)}{\partial \hat{a}_P^K(\mathbf{x}_K)} \cdot \frac{\partial \hat{a}_P^K(\mathbf{x}_K)}{\partial x_K^d} + \sum_{i,j} \frac{\partial j(\mathbf{X}_K)}{\partial [\mathbf{C}_{KK}^n]_{ij}} \cdot \frac{\partial [\mathbf{C}_{KK}^n]_{ij}}{\partial x_K^d} \end{aligned} \quad (\text{A.2})$$

where  $[\mathbf{C}_{KK}^n]_{ij}$  is the  $ij$ -th entry of the noise covariance matrix. The derivatives  $\frac{\partial j(\mathbf{X}_K)}{\partial \hat{a}_p^K(\mathbf{x}_K)}$  and  $\frac{\partial j(\mathbf{X}_K)}{\partial [\mathbf{C}_{KK}^n]_{ij}}$  can be simply obtained from (A.1). The derivatives of the spatial gains are themselves Gaussian processes with the following parameters

$$\frac{\partial \hat{a}_p^K(\mathbf{x}_K)}{\partial x_K^d} \sim \mathcal{GP} \left( \frac{\partial m^{a_p}(\mathbf{x}_K)}{\partial x_K^d}, \frac{\partial C^{a_p}(\mathbf{x}_K, \mathbf{x}'_K)}{\partial x_K^d \partial x'^d_K} \right). \quad (\text{A.3})$$

The vector containing the spatial gains at the sensor positions and the derivatives of the spatial gain at the new sensor position,  $\mathbf{v}_p = [\hat{a}_p^1(\mathbf{x}_1), \hat{a}_p^2(\mathbf{x}_2), \dots, \hat{a}_p^M(\mathbf{x}_M), \frac{\partial \hat{a}_p^M(\mathbf{x}_M)}{\partial x_M^1}, \frac{\partial \hat{a}_p^M(\mathbf{x}_M)}{\partial x_M^2}, \dots, \frac{\partial \hat{a}_p^M(\mathbf{x}_M)}{\partial x_M^D}]^T$ , possesses a joint Gaussian distribution, for  $p = 1, 2, \dots, P$ . The covariance between the spatial gain at the point  $\mathbf{x}_K$  ( $k = 1, 2, \dots, K$ ) and the derivatives of spatial gain at the point  $\mathbf{x}_K$  is given by

$$\text{Cov} \left( \hat{a}_p^k(\mathbf{x}_k), \frac{\partial \hat{a}_p^K(\mathbf{x}_K)}{\partial x_K^d} \right) = \frac{\partial C^{a_p}(\mathbf{x}_k, \mathbf{x}_K)}{\partial x_K^d}, \quad (\text{A.4})$$

and the covariance between derivatives with respect to different dimensions  $d_1, d_2 \in \{1, 2, \dots, D\}$  is given by

$$\text{Cov} \left( \frac{\partial \hat{a}_p^K(\mathbf{x}_K)}{\partial x_K^{d_1}}, \frac{\partial \hat{a}_p^K(\mathbf{x}_K)}{\partial x_K^{d_2}} \right) = \frac{\partial C^{a_p}(\mathbf{x}'_K, \mathbf{x}_K)}{\partial x'^{d_1}_K \partial x_K^{d_2}} \Big|_{\mathbf{x}'_K = \mathbf{x}_K}. \quad (\text{A.5})$$

Now, through the distribution of the random vectors  $\mathbf{v}_p$ , their samples can be generated for  $p = 1, 2, \dots, P$ . Subsequently, utilizing (A.2), the samples of  $\nabla_{\mathbf{x}_K} j(\mathbf{X}_{K-1}, \mathbf{x}_K)$  can be obtained.

## References

- [1] Y. Chen, C.-N. Chuah, Q. Zhao, Sensor placement for maximizing lifetime per unit cost in wireless sensor networks, in: IEEE Mil. Commun. Conf., IEEE, 2005, pp. 1097–1102.
- [2] E. S. Biagioni, G. Sasaki, Wireless sensor placement for reliable and efficient data collection, in: Proc. 36th Annu. Hawaii Int. Conf. Syst. Sci., IEEE, 2003, pp. 10–pp.

- [3] W. Ostachowicz, R. Soman, P. Malinowski, Optimization of sensor placement for structural health monitoring: A review, *Struct. Health Monit.* 18 (3) (2019) 963–988.
- [4] J. S. Abel, Optimal sensor placement for passive source localization, in: *IEEE Int. Conf. Acoust. Speech Signal Process.*, IEEE, 1990, pp. 2927–2930.
- [5] N. Sahu, L. Wu, P. Babu, B. S. MR, B. Ottersten, Optimal sensor placement for source localization: A unified ADMM approach, *IEEE Trans. Veh. Technol.* 71 (4) (2022) 4359–4372.
- [6] P. Comon, C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.
- [7] J.-F. Cardoso, B. H. Laheld, Equivariant adaptive source separation, *IEEE Trans. Signal Process.* 44 (12) (1996) 3017–3030.
- [8] S. Choi, A. Cichocki, S. Amari, Equivariant nonstationary source separation, *Neural Netw.* 15 (1) (2002) 121–130.
- [9] J. R. Hopgood, P. J. Rayner, P. W. Yuen, The effect of sensor placement in blind source separation, in: *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, IEEE, 2001, pp. 95–98.
- [10] Q. Wang, S. Y. Low, Z. Li, K.-f. C. Yiu, Sensor placement optimization of blind source separation in a wireless acoustic sensor network via hybrid descent methods, *Appl. Acoust.* 188 (2022) 108509.
- [11] H. Cecotti, B. Rivet, M. Congedo, C. Jutten, O. Bertrand, E. Maby, J. Matout, A robust sensor-selection method for P300 brain–computer interfaces, *J. Neural Eng.* 8 (1) (2011) 016001.
- [12] H. Zhang, R. Ayoub, S. Sundaram, Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms, *Automatica* 78 (2017) 202–210.
- [13] S. Xu, K. Doğançay, Optimal sensor placement for 3-d angle-of-arrival target localization, *IEEE Trans. Aerosp. Electron. Syst.* 53 (3) (2017) 1196–1211.
- [14] S. Xu, L. Wu, K. Doğançay, M. Alae-Kerahroodi, A hybrid approach to optimal toa-sensor placement with fixed shared sensors for simultaneous multi-target localization, *IEEE Trans. Signal Process.* 70 (2022) 1197–1212.
- [15] N. Sahu, L. Wu, P. Babu, B. S. MR, B. Ottersten, Optimal sensor placement for source localization: A unified admm approach, *IEEE Trans. Veh. Technol.* 71 (4) (2022) 4359–4372.

- [16] F. Ghayem, B. Rivet, C. Jutten, R. C. Farias, Optimal sensor placement for signal extraction, in: Proc. IEEE Int. Conf. Acoust. Speech Signal Process., IEEE, 2019, pp. 4978–4982.
- [17] F. Ghayem, B. Rivet, R. C. Farias, C. Jutten, Robust sensor placement for signal extraction, IEEE Trans. Signal Process. 69 (2021) 4513–4528.
- [18] M. Sadeghi, B. Rivet, M. Babaie-Zadeh, Optimal sensor placement for source separation with noisy measurements, in: Proc. 31st Eur. Signal Process. Conf., IEEE, 2023, pp. 1664–1668.
- [19] S. Makeig, T.-P. Jung, A. J. Bell, D. Ghahremani, T. J. Sejnowski, Blind separation of auditory event-related brain responses into independent components, Proc. Natl. Acad. Sci. U.S.A. 94 (20) (1997) 10979–10984.
- [20] N. Xu, X. Gao, B. Hong, X. Miao, S. Gao, F. Yang, BCI competition 2003-data set IIB: enhancing p300 wave detection using ICA-based subspace projections for BCI applications, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1067–1072.
- [21] S. Shahbazpanahi, A. Gershman, Z.-Q. Luo, K. M. Wong, Robust adaptive beamforming for general-rank signal models, IEEE Trans. Signal Process. 51 (9) (2003) 2257–2269.
- [22] C. K. Williams, C. E. Rasmussen, Gaussian processes for machine learning, Vol. 2, MIT press Cambridge, MA, 2006.
- [23] R. Ge, F. Huang, C. Jin, Y. Yuan, Escaping from saddle points—online stochastic gradient for tensor decomposition, in: Proc. COLT, PMLR, 2015, pp. 797–842.
- [24] A. Hyvärinen, J. Karhunen, E. Oja, Independent component analysis, John Wiley & Sons, 2001.
- [25] M. C. Shewry, H. P. Wynn, Maximum entropy sampling, J. Appl. Stat. 14 (2) (1987) 165–170.
- [26] W. F. Caselton, J. V. Zidek, Optimal monitoring network designs, Stat. Probab. Lett. 2 (4) (1984) 223–227.
- [27] A. Krause, A. Singh, C. Guestrin, Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies., J. Mach. Learn. Res. 9 (2) (2008).
- [28] M.-H. Lee, O.-Y. Kwon, Y.-J. Kim, H.-K. Kim, Y.-E. Lee, J. Williamson, S. Fazli, S.-W. Lee, EEG dataset and OpenBMI toolbox for three BCI paradigms: An investigation into BCI illiteracy, GigaScience 8 (5) (2019) giz002.

- [29] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, T. M. Vaughan, Brain-computer interfaces for communication and control., *Clinical Neurophysiology* 113 (6) (2002) 767–791.
- [30] B. Rivet, A. Souloumiac, V. Attina, G. Gibert, xDAWN algorithm to enhance evoked potentials: application to brain–computer interface, *IEEE. Trans. Biomed. Eng.* 56 (8) (2009) 2035–2043.
- [31] N. Cressie, *Statistics for spatial data*, John Wiley & Sons, 1991.
- [32] U. Hoffmann, J.-M. Vesin, T. Ebrahimi, K. Diserens, An efficient p300-based brain–computer interface for disabled subjects, *J. Neurosci. Methods* 167 (1) (2008) 115–125.
- [33] D. J. MacKay, Bayesian interpolation, *Neural Comput.* 4 (3) (1992) 415–447.
- [34] C. Guger, S. Daban, E. Sellers, C. Holzner, G. Krausz, R. Carabalona, F. Gramatica, G. Edlinger, How many people are able to control a P300-based brain-computer interface (BCI)?, *Neurosci. Lett.* 462 (1) (2009) 94–98.
- [35] D. J. Krusienski, E. W. Sellers, F. Cabestaing, S. Bayouhd, D. J. McFarland, T. M. Vaughan, J. R. Wolpaw, A comparison of classification techniques for the P300 speller, *J. Neural Eng.* 3 (4) (2006) 299.